Stat 414 – Day 11
Random Slopes cont., Centering (4.6, 5.2)

Last Time: Random slopes (Ch. 5)

- $Y_{ij} = \beta_{00} + u_{0j} + (\beta_{10} + u_{1j})x_{ij} + \epsilon_{ij}$
  - Effect of x_{ij} differs among Level 2 groups (e.g., schools)
  - Adds variance component and covariance between slopes and intercepts
- Var(Y_{ij}) = $\tau_0^2 + 2\tau_{01}x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2$
  - Results impacted by choice of scaling (origin) of *x* variables
  - Smallest when $x = -\tau_{01}/\tau_1^2$ (correlation vs. covariance)
  - Expect 95% of slopes within $\beta_{10} \pm 2\tau_1$
- Cov of two individuals in same group: $\tau_0^2 + \tau_{01}(x_{aj} + x_{bj}) + \tau_1^2(x_{aj}x_{bj})$
  - No simple ICC (depends on *x*)
- Explaining level 2 variance
  - Consider intercepts, slopes of "outcomes"

$\text{corr} = \frac{\text{cov}(x,y)}{\text{SD}(x)\,\text{SD}(y)}$

Keep in mind: With random slopes,

Corr(Y_{aj}, Y_{bj}) = Cov(Y_{aj}, Y_{bj})/SD(Y_{aj}) SD(Y_{bj})

$$= \frac{\tau_0^2+\tau_{01}(x_{aj}+x_{bj})+\tau_1^2(x_{aj}x_{bj})}{\sqrt{(\tau_0^2+2\tau_{01}x_{aj}+\tau_1^2x_{aj}^2+\sigma^2)(\tau_0^2+2\tau_{01}x_{bj}+\tau_1^2x_{bj}^2+\sigma^2)}}$$

$\hat{\tau}_{01} = -.86 \times .5296 \times .1386$
$= -.064$

For Netherlands data:

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 school   (Intercept) 0.28050  0.5296
          gevocab     0.01922  0.1386   -0.86
 Residual             3.66613  1.9147
```

getVarCov(model1b, type = "marginal")
cov2cor(cov$'767')
cov2cor(cov[[1]])

$x_a = 1.7$ $x_b = 3.1$

$.2805 + (-.064)(4.8) + .01922(1.7)(3.1) = .075$

$$\frac{.075}{\sqrt{.2805 + 2(-.064)(1.7) + (.01922)(1.7^2) + .3666} \times \sqrt{.2805 + 2(-.0641)(3.1) + (.01922)(3.1^2) + .3666}}$$
$= .021$

$x_a = 1.7$ $x_b = 9.3$ $\Rightarrow -.027$

Not a simple intraclass correlation coefficient with random slopes.

Example 1: Day 10 Example 2 cont. (lmer)

```
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 school   (Intercept) 3.6391814 1.90766
          gevocab     0.2605131 0.51040  -0.08
          age         0.0005558 0.02357  -0.69 -0.66
 Residual             3.6076583 1.89938

Fixed effects: geread ~ gevocab + age
                Value Std.Error    DF
(Intercept)  2.9614055 0.4151887 10158
gevocab      0.5191496 0.0143563 10158
age         -0.0088390 0.0038396 10158
```

```
Random effects:
 Groups   Name        Variance   Std.Dev. Corr
 school   (Intercept) 0.10238213 0.319972
          cgevocab    0.01906890 0.138090  0.53
          cage        0.00002486 0.004986 -0.28 -0.96
 Residual             3.66396826 1.914149

Fixed effects:
             Estimate  Std. Error t value
(Intercept)  4.343884  0.032704  132.825
cgevocab     0.519277  0.014361   36.159
cage        -0.008882  0.003822   -2.324
```

when gevocab = mean
age = mean age

c gevocab = gevocab − mean(gevocab)

Definition: ***Grand mean centering*** is subtracting off the overall mean, $y_{ij} - \bar{y}$. Benefits include interpretation of intercepts, reduce collinearity in "product terms." Can also be especially helpful with strong slope/intercept covariance. "When working with random coefficient models, it is best that ***all*** independent variables be centered.

> ***Group mean centering*** is subtracting off the group mean, $y_{ij} - \bar{y}_j$. This is not equivalent and impacts interpretation, but is less common in practice.

(a) Can we include an interaction between age and gevocab in the two cases? How do we interpret this interaction?

(b) How does centering *gevocab* impact our earlier analysis?

(c) What if don't want the random slopes to be correlated?
(age | school) + (gevocab | school)

Example 2: Recall the data predicting *PctBush* with counties clustered within states. We want to consider the percentage of families in the county in a traditional nuclear form as predictor.

(a) What is the overall mean *PctFamily*? 19.036 $\bar{y}$

(b) What are the means of *PctFamily* for the first four states?

AL 17.519 AR 17.801 AZ 18.247 CA 21.088 $\bar{y}_j$

(c) What would PctFamily be if grand-mean centered? Group-mean centered?

state_po	county	PctBush	Income	PctFamily	Grand mean centered	Group mean centered
AL	Autauga	75.67352	48.458	25.3	25.3 − 19.036 = 6.26	25.3 − 17.519 = 7.78
AR	Boone	66.27191	34.974	19.6	19.6 − 19.036 = 0.564	19.6 − 17.801
CA	Orange	59.80009	64.611	26.1	26.1 − 19.036 = 7.06	26.1 − 21.088

(d) In a simple model with a fixed slope for PctFamily, how do the slope and intercept compare depending on which version you use? Why?

	Intercept	Slope
Raw data no centering	40.21	.897
Grand mean	57.29	.897
Group mean	57.01	.898

same slope but different intercept

similar to grand mean centering

(e) In a model with a random slope for PctFamily, how do the slope and intercept and slope/intercept correlation values compare for the different versions? What about the prediction for Autauga?

	Intercept	Slope	Corr	$\hat{y}$
Raw data	39.77	.938	−.84	82.97
Grand mean	57.64	.938	0.10	82.97
Group mean	56.97	.9224	−.04	82.88 ← not equivalent

(f) Which models are equivalent in the predicted values? Why?
raw data (no centering) & grand mean (shifting)
group mean centering creates a different variable

In considering Level 2 variables to add to the model, an "especially important type" of "contextual variable" is the group mean.

(g) Include pctfamilyavg in the model, along with the interaction with PctFam. high VIF

(h) Include StatePctFam, grand-mean centered, in the model and compare interpretations depending on how the Level 1 PctFam variable is centered.

	Intercept	PctFam	StatePctFam		
Raw data	17.4	0.89	1.22		
Both Grand mean	57.5	0.89	1.22	change in PctBush for 2 counties w/ same PctFam but StatePctFam differs by 1	$y, x, \bar{x}$
L1: Group mean/ L2: Grand mean	57.5	0.89*	2.11	Change in mean PctBush when mean State Pct Family differs by 1	$\bar{y}, \bar{x}$

Note: Testing the group level coefficient (coefficient of the group mean) is referred to as the Hausman specification test in econometrics.

(i) What does a positive coefficient for the group mean tell you?
w/ grand mean: coefficient $\bar{x}_j$ represents additional increase in response w/ higher $\bar{x}_j$ at Level 2

(j) What do you notice about the sum of the slope coefficients in row 2 and the slope coefficients in row 3?
slope at Level 2 (between group) 2.11 = .89 + 1.22
within group additional contribution

(k) How do these results compare to the "within group" regression and "between group" regression?
2.45 "effect" w/ one-unit increase in student IQ w/ school mean IQ "fixed"
1.28 is increase in slope for school level effect

(l) How do you summarize and evaluate the significance of the "contextual effect"?
the contextual effect here is the 1.28

language

Example 3: Reconsider the Netherlands data on language score based on verbal IQ from the text (e.g., Section 4.6)

(a) Fit the model with both IQ and mean IQ. What is the slope of the "within group" regression?
2.45

(b) What is the additional contribution to langPOST from the class mean verbal IQ? Is this statistically significant? 1.28 $t = 4.906$
=> statistically significant

(c) What if we put both grand-mean centered IQ and grand-mean centered mean IQ in the model; how do we interpret the coefficient of mean IQ in this model?

1.28 additional contribution

sometimes called the "contextual effect" - it's what the school contributes to the prediction above and beyond the individual student prediction (two students with same IQ but below to schools that are 1 apart on school average IQ)

(d) How does the coefficient of mean IQ change when we use the group-mean centered verbal IQ?

$3.74 = 1.28 + 2.45$

with group mean centering, slope of mean IQ is 3.75. You might prefer this information if your focus in on relationships at the group level (e.g., funding decisions); effect of increase in school average on the response average (see question e)

(e) How does this coefficient compare to the slope of mean IQ in the aggregated data set?

Same idea, similar

If regress 211 language means (one for each school) on 211 IQ means (one for each school), get a slope of 3.78

(f) Interpret the slope/intercept correlation in the multilevel model. (first make IQ have random slopes)

random slopes for IQ centered

as intercepts increase, slopes decrease

(g) What about including SES in the model? (see day 12)